Towards More Accurate and Efficient Beamformed Radio Interferometry Imaging

MASTER THESIS

Author:
Matthieu Simeoni, EPFL

Supervisors:
Dr. Paul Hurley, IBM Research, Zürich
Prof. Victor Panaretos, SMAT, EPFL
Prof. Martin Vetterli, LCAV, EPFL

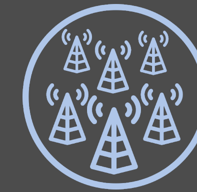
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Astronomical Data Deluge...

- Modern radio interferometers combine many small antennas together in a **phased array.**
- Data generated by such these instruments is **enormous** !
- Need to **reduce** the amount of data sent to the central processor.

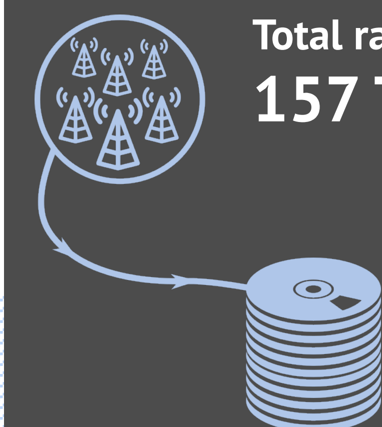
The SKA1 LOW in brief

Location:
Australia



~130,000 antennas
100 stations

Maximum distance between two stations :



Total raw output :
157 Tb/s

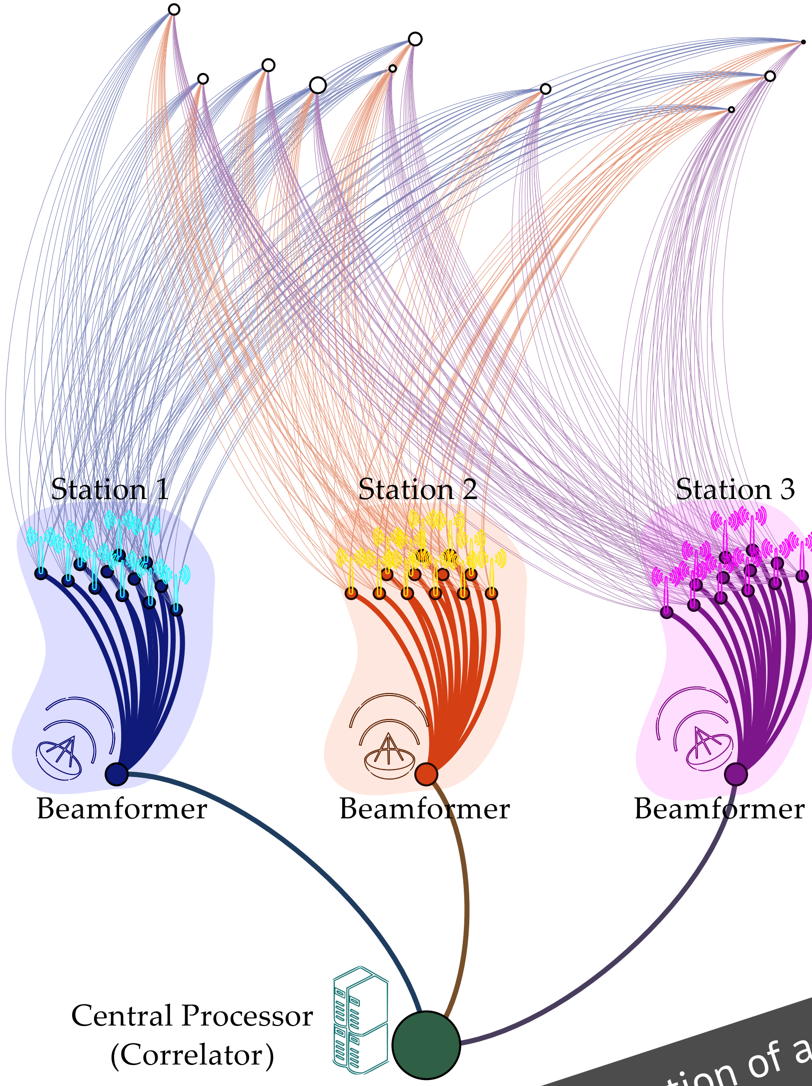
Enough to fill up **350,000 DVDs** per second

Hierarchical Designs

- Antennas are grouped together in **stations.**
- Data is **beamformed** at the station level before being sent to the **central processor**:

$$\omega_i : \begin{cases} \mathbb{C}^L \to \mathbb{C}, \\ \mathbf{x}_i(t) \mapsto y_i(t) = \mathbf{w}_i^H \mathbf{x}_i(t). \end{cases}$$

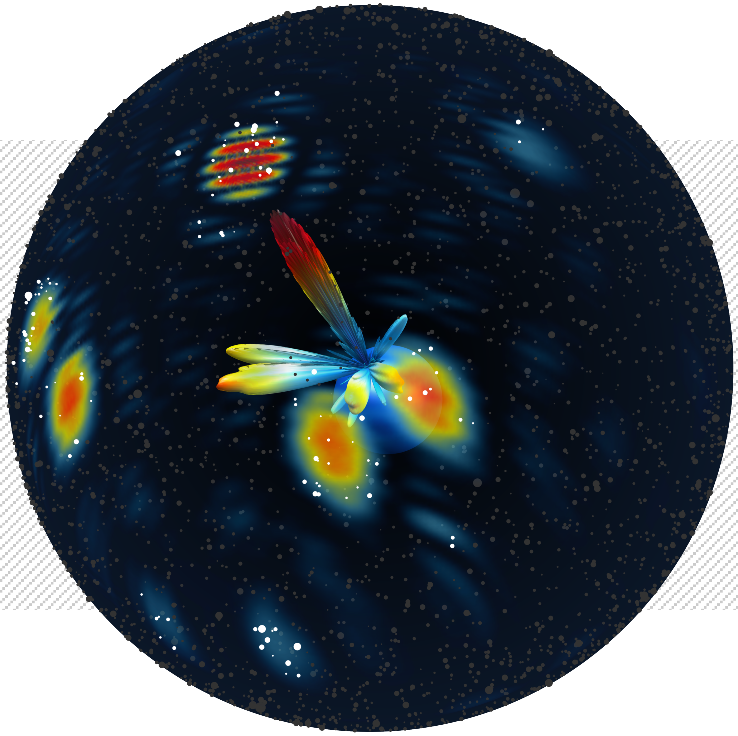
- Strategy currently deployed in **LOFAR.**
- But **how** should we beamform?



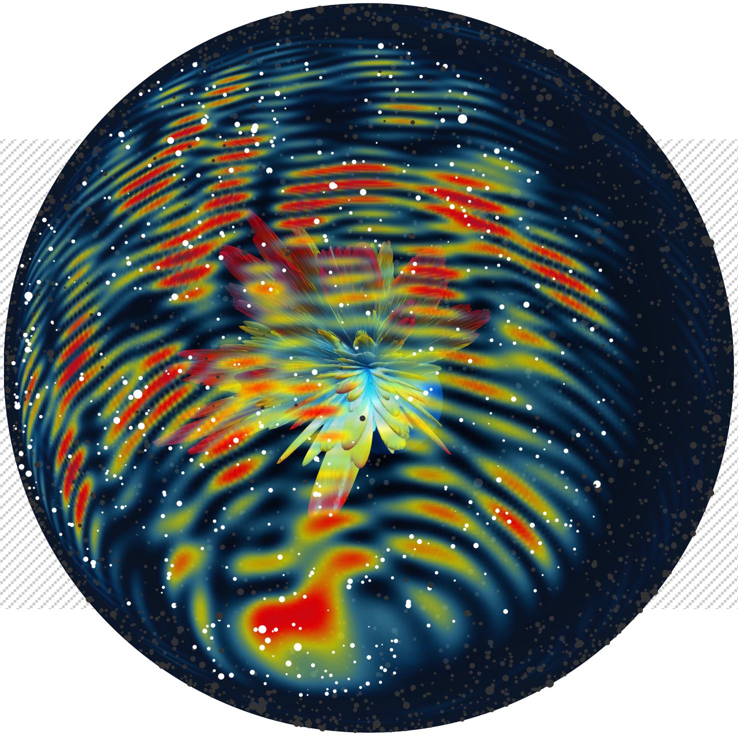
Schematic representation of a hierarchical design for modern radio telescopes

Generalized Beamforming

Matched Beamforming vs. **Randomized Beamforming**



Maximize signal power



Maximize sky coverage

CURRENT IMAGING PIPELINE NOT FLEXIBLE ENOUGH !

Table of Contents

Classical Data Model (No Beamforming)

S1. Sources lie on an hypothetical sphere (celestial sphere)

S2. Signals can be viewed parallel (far field)

S3. Narrow band signals

S4. Signals from different positions in the sky are uncorrelated

MODELING ASSUMPTIONS

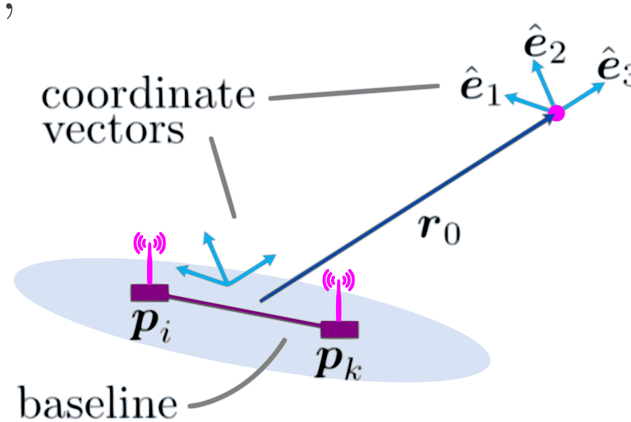
- Signals measured by each antenna are correlated:

$$V_{i,k} := \mathbb{E}[x_i(t)x_k^*(t)] = \iint_{\mathbb{S}^2} I(\mathbf{r})e^{-j2\pi\langle \mathbf{r}, \frac{\mathbf{p}_i - \mathbf{p}_k}{\lambda_0}\rangle} d\mathbf{r}.$$

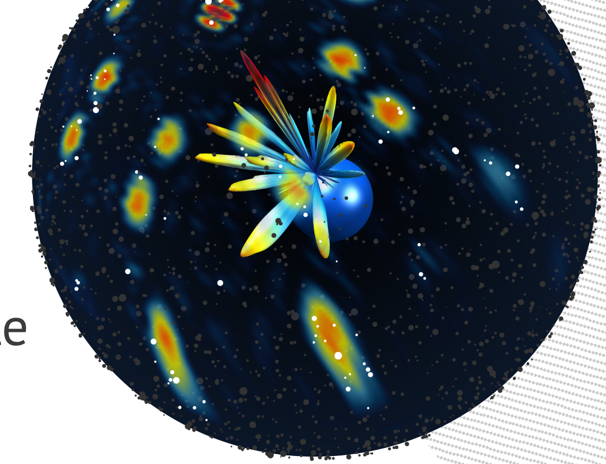
- For small field of views, almost 2D Fourier transform

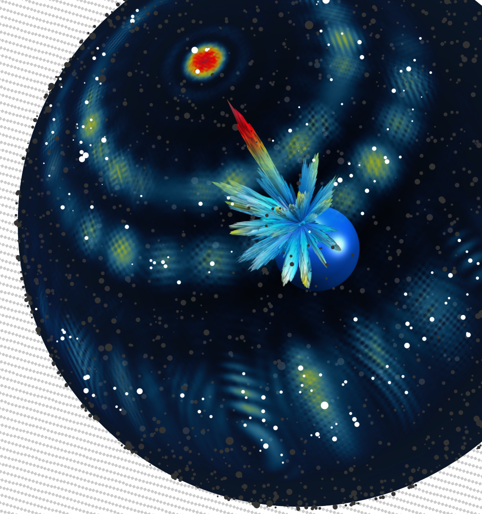
$$V_{i,k} \simeq e^{-j2\pi w_{i,k}} \iint_{K\subset\mathbb{R}^2} I(l,m)e^{-j2\pi(u_{i,k}l + v_{i,k}m)},$$

- Visibilities can be seen as Fourier samples.



Classical Data Model (With Beamforming)

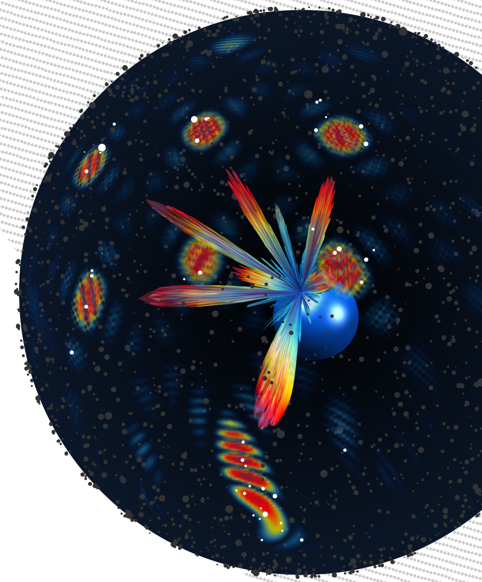




- For a sky composed of a single source,

$$\mathbb{E}[y_i(t)y_i^*(t)] = |\mathbf{w}_i^H \mathbf{a}_i(\mathbf{r}_q)|^2 \sigma_q^2 \ + \ \|\mathbf{w}_i\|^2 \sigma_n^2.$$

- $b_i(\mathbf{r}) = \mathbf{w}_i^H \mathbf{a}_i(\mathbf{r})$ is the **beamshape** of station *i*.

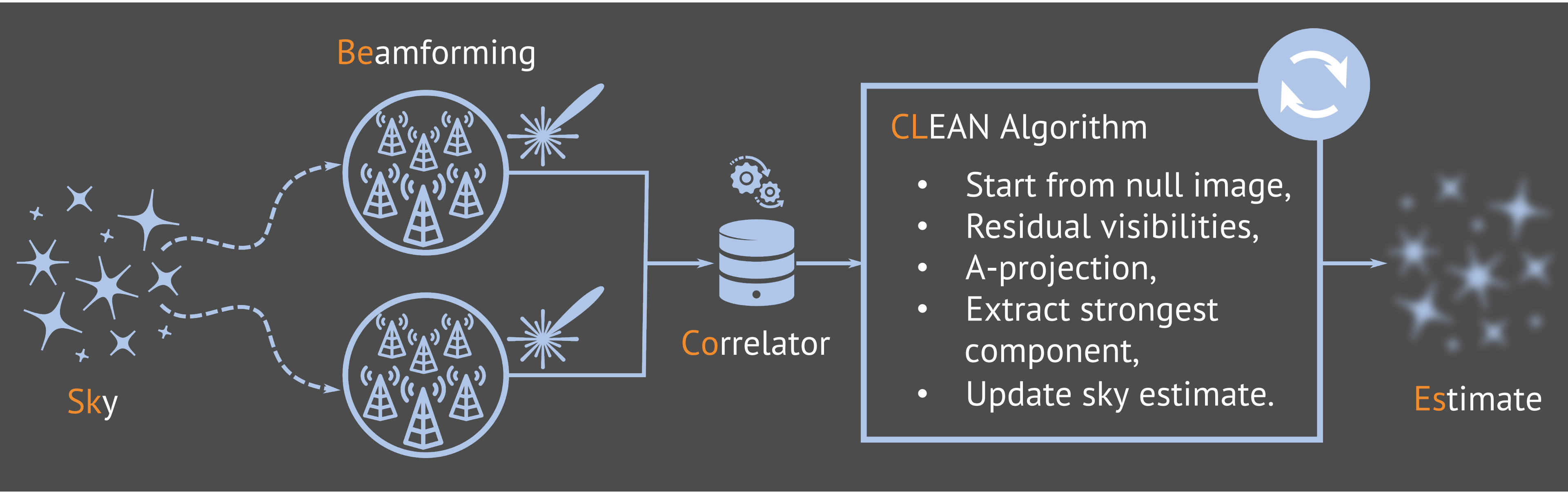


Virtual Antenna Assumption:

- Hence, the data model for beamformed data is given by

$$V_{i,k} = \iint_{K \subset \mathbb{R}^2} I(l,m) b_i(l,m) b_k^*(l,m) \mathcal{W}_{i,k}(l,m) e^{-j2\pi(u_{i,k}l + v_{i,k}m)} dl dm$$

Current Imaging Pipeline



The CLEAN Algorithm

Gradient descent finds a local minimum of a function by taking steps in the opposite direction of the gradient.

REMINDER

- Find solution to linear system

$$\boldsymbol{\mathcal{V}} = \mathcal{A}\boldsymbol{I}.$$

- **CLEAN** produces "sparse" sky estimates

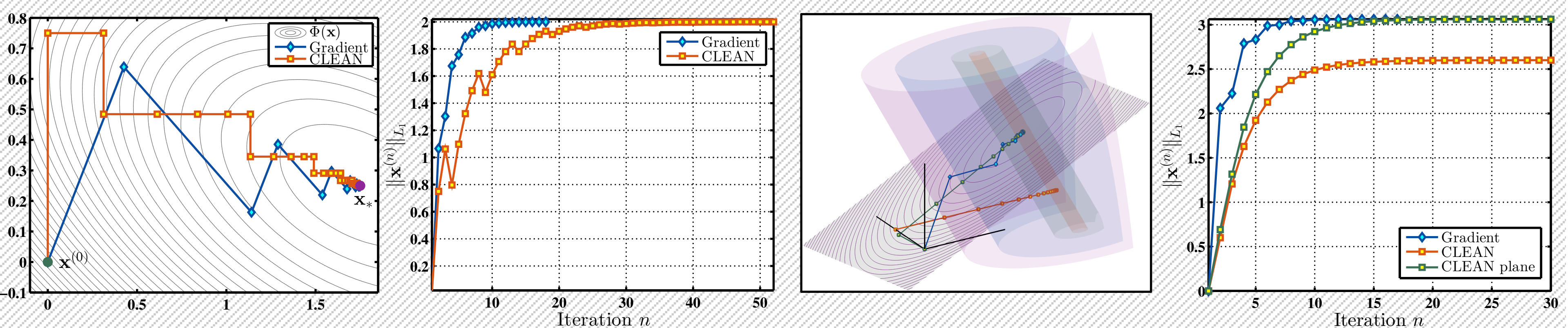
$$\hat{\boldsymbol{I}}^{(n+1)} = \hat{\boldsymbol{I}}^{(n)} + \tau\Psi\mathcal{A}^{H}(\boldsymbol{\mathcal{V}} - \mathcal{A}\hat{\boldsymbol{I}}^{(n)}).$$

- Nonlinear, and very sensitive on the choice of τ.

- Can be seen as an approximate **gradient descent**.

Typically > 10,000 iterations

CLEAN as Gradient Descent



The A-projection Algorithm

Cross-Beamshapes
+w-term

Sampling kernel

- Recall the measurement equation

$$V_{i,k} = \boxed{\iint_{K\subset\mathbb{R}^2} I(l,m)\boxed{b_i(l,m)b_k^*(l,m)\mathcal{W}_{i,k}(l,m)}\, e^{-j2\pi\boxed{(u_{i,k}l+v_{i,k}m)}}\,dl\,dm}$$

- Hence,

Fourier transform

$$\boldsymbol{\mathcal{V}} = \mathcal{A}\boldsymbol{I} = \boxed{\mathcal{S}}\;\boxed{\mathcal{F}}\;\boxed{\mathcal{B}}\;\boldsymbol{I}.$$

The A-projection Algorithm

Cross-Beamshapes +w-term

Sampling kernel

- Recall the measurement equation

$$V_{i,k} = \boxed{\iint_{K \subset \mathbb{R}^2} I(l,m) \boxed{b_i(l,m) b_k^*(l,m) \mathcal{W}_{i,k}(l,m)} e^{-j2\pi \boxed{(u_{i,k} l + v_{i,k} m)}} dl dm}$$

- Hence,

Fourier transform

$$\boldsymbol{\mathcal{V}} = \mathcal{A}\boldsymbol{I} = \boxed{\mathcal{S}}\ \boxed{\mathcal{F}}\ \boxed{\mathcal{B}}\ \boldsymbol{I}.$$

- The **A-projection** algorithm allows fast multiplication by $\mathcal{A}$ or $\mathcal{A}^H$

$$\mathcal{A}\boldsymbol{I} = \mathcal{S}\mathcal{F}\mathcal{B}\boldsymbol{I} = \mathcal{S}\left(\hat{\mathcal{B}} * \mathcal{F}\boldsymbol{I}\right).$$

Convolution theorem

The A-projection Algorithm

Cross-Beamshapes +w-term

Sampling kernel

- Recall the measurement equation

$$V_{i,k} = \boxed{\iint_{K\subset\mathbb{R}^2} I(l,m)\boxed{b_i(l,m)b_k^*(l,m)\mathcal{W}_{i,k}(l,m)}\,e^{-j2\pi\boxed{(u_{i,k}l+v_{i,k}m)}}dldm}$$

- Hence,

Fourier transform

$$\boldsymbol{\mathcal{V}} = \mathcal{A}\boldsymbol{I} = \boxed{\mathcal{S}}\,\boxed{\mathcal{F}}\,\boxed{\mathcal{B}}\,\boldsymbol{I}.$$

- The **A-projection** algorithm allows fast multiplication by $\mathcal{A}$ or $\mathcal{A}^H$

$$\mathcal{A}\boldsymbol{I} = \mathcal{S}\mathcal{F}\mathcal{B}\boldsymbol{I} = \mathcal{S}\left(\hat{\mathcal{B}} * \mathcal{F}\boldsymbol{I}\right).$$

Convolution theorem

- A-projection can also be used to approximation the **pseudoinverse**

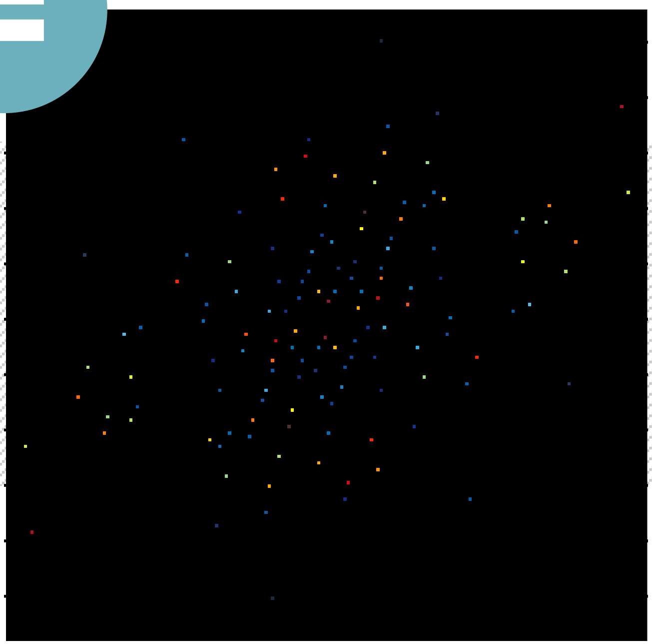
$$\left(\mathcal{A}^H\mathcal{A}\right)^{-1}\mathcal{A}^H\boldsymbol{\mathcal{V}} \simeq \left(\mathcal{B}^H\mathcal{B}\right)^{-1}\mathcal{A}^H\boldsymbol{\mathcal{V}}.$$

A-projection and Fourier

Equivalent Telescope Assumption is a fallacy !

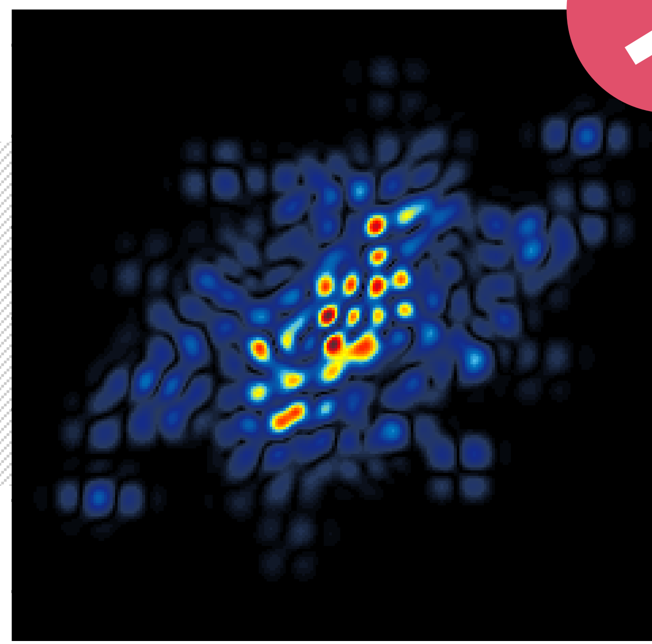
- Interpreted at the continuous level, applying A-projection yields

$$\mathcal{F}^{-1}\left\{\sum_{i,k}\mathcal{F}\{b_i^* b_k\} * \tilde{V}_{i,k}\right\}$$



Spectrum before A-projection

#SPECTRUM = #VISIBILITIES



Spectrum after A-projection

#SPECTRUM > #VISIBILITIES

The Natural Measurement Equation

- Direct correlation computation between two beamformed outputs yields

$$V_{i,k} = \iint_{\mathbb{S}^2} I(\boldsymbol{r}) b_i(\boldsymbol{r}) b_k^*(\boldsymbol{r}) d\boldsymbol{r}.$$

No Fourier Kernel !

- No apparent link with Fourier, so stay on the sphere.

- Geometric interpretation

$$V_{i,k} = \langle I, b_i^* b_k \rangle = \langle I, \beta_{i,k} \rangle.$$

  - Inner product with periodic functions,
  - Functions given by telescope layout and beamformer.

The Gram-Schmidt Imager

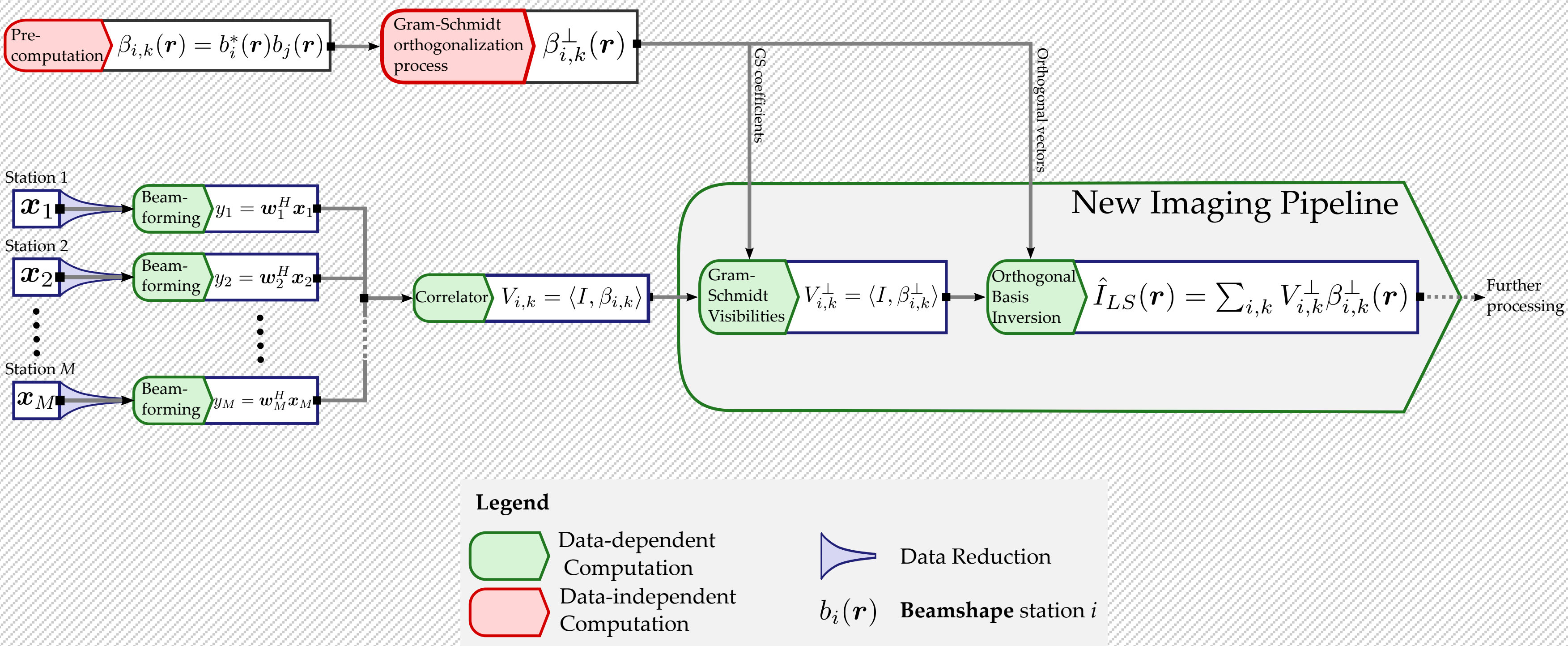
1. Reconstruction on the sphere
2. Valid at the continuous lvl
3. Linear in the data
4. No gridding/FFT
5. Direct Solver

ADVANTAGES

- **Idea:** Orthogonalize the instrument, and modify the visibilities accordingly
- Compute least squares estimate as

$$\hat{I}_{LS}(\boldsymbol{r}) = \sum_{i,k} V_{i,k}^{\perp} \beta_{i,k}^{\perp}(\boldsymbol{r}).$$

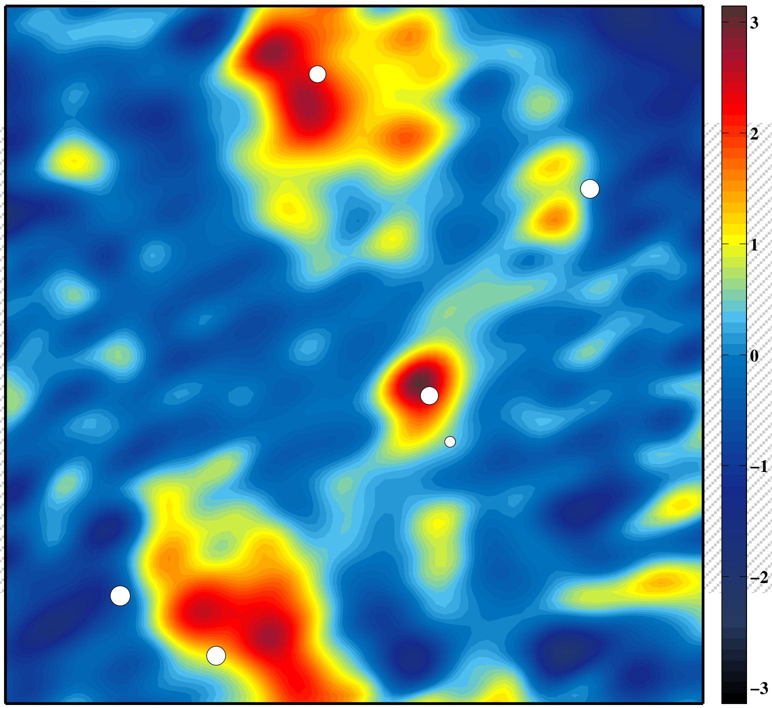
- For efficiency and stability, use **QR-factorization**



Comparison with A-projection

Case: Matched Beamforming

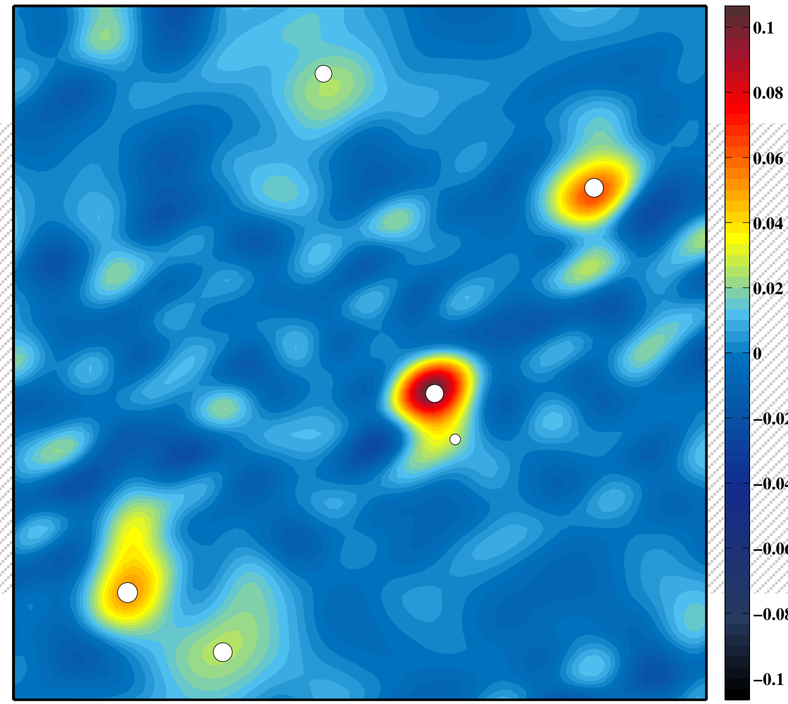
A-projection Estimate



Sources not well resolved, severely polluted by artifacts

VS.

Gram-Schmidt Estimate



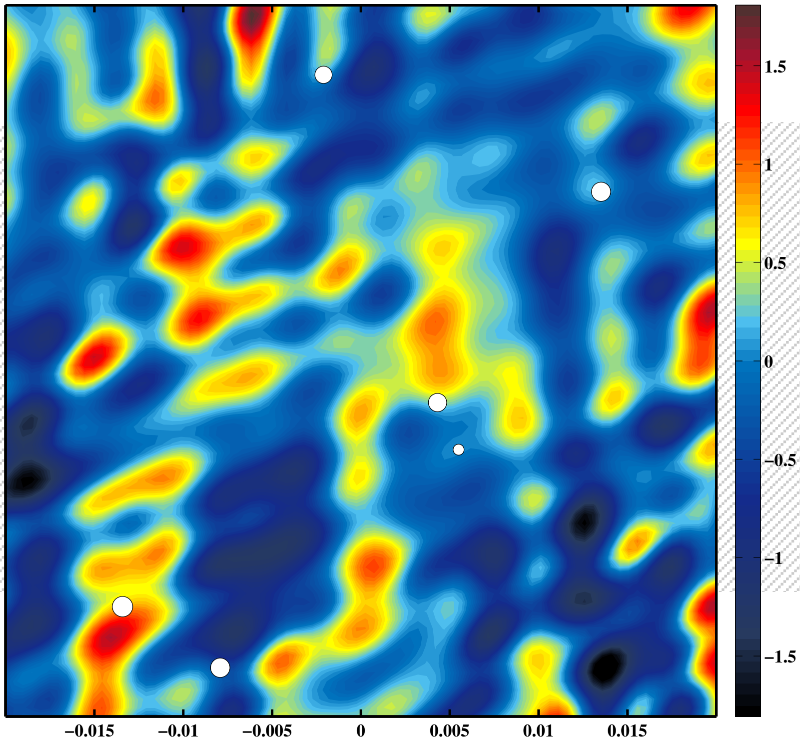
Almost all sources are resolved, small artifacts

GRAM-SCHMIDT ESTIMATE MORE ACCURATE

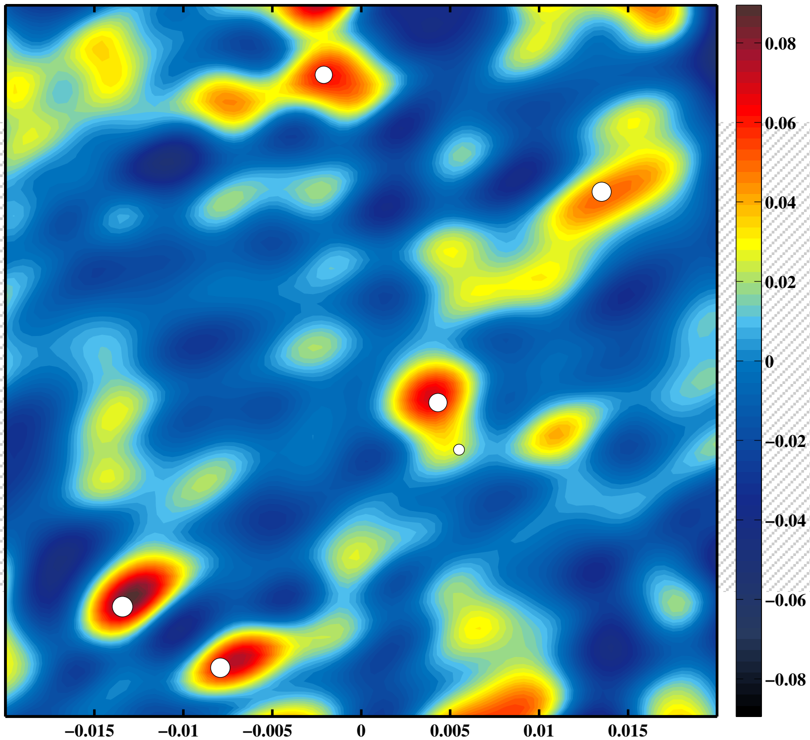
Comparison with A-projection

Case: Randomized Beamforming

A-projection Estimate vs. **Gram-Schmidt Estimate**



Estimate completely unreliable



Almost all sources are resolved, small artifacts

GRAM-SCHMIDT IMAGER MORE FLEXIBLE

Robustness to Noise

- GS involves the following steps

for $i = 2$ **to** J **do**

$\tilde{\boldsymbol{\beta}}_i \leftarrow \boldsymbol{\beta}_i - \sum_{k=1}^{j-1} \langle \boldsymbol{\beta}_i, \boldsymbol{\beta}_k^{\perp} \rangle \boldsymbol{\beta}_k^{\perp}$;

if $\|\tilde{\boldsymbol{\beta}}_i\|_2 \neq 0$ **then**

$j \leftarrow j + 1$;

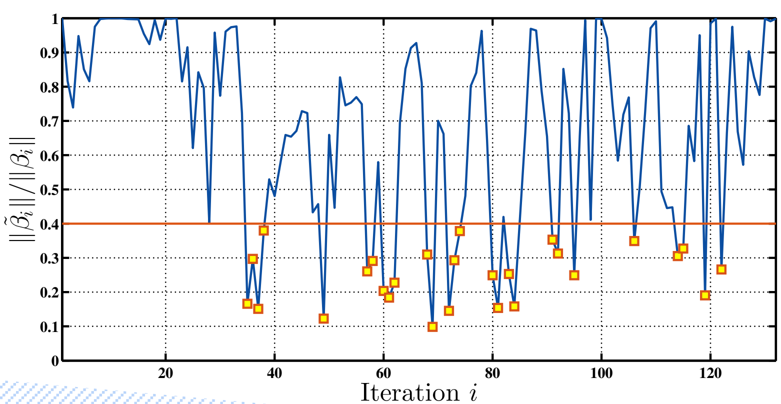
$\boldsymbol{\beta}_j^{\perp} \leftarrow \tilde{\boldsymbol{\beta}}_i / \|\tilde{\boldsymbol{\beta}}_i\|_2$;

$V_j^{\perp} \leftarrow \left(V_i - \sum_{k=1}^{j-1} \langle \boldsymbol{\beta}_i, \boldsymbol{\beta}_k^{\perp} \rangle V_k^{\perp} \right) \boxed{/ \|\tilde{\boldsymbol{\beta}}_i\|}$;

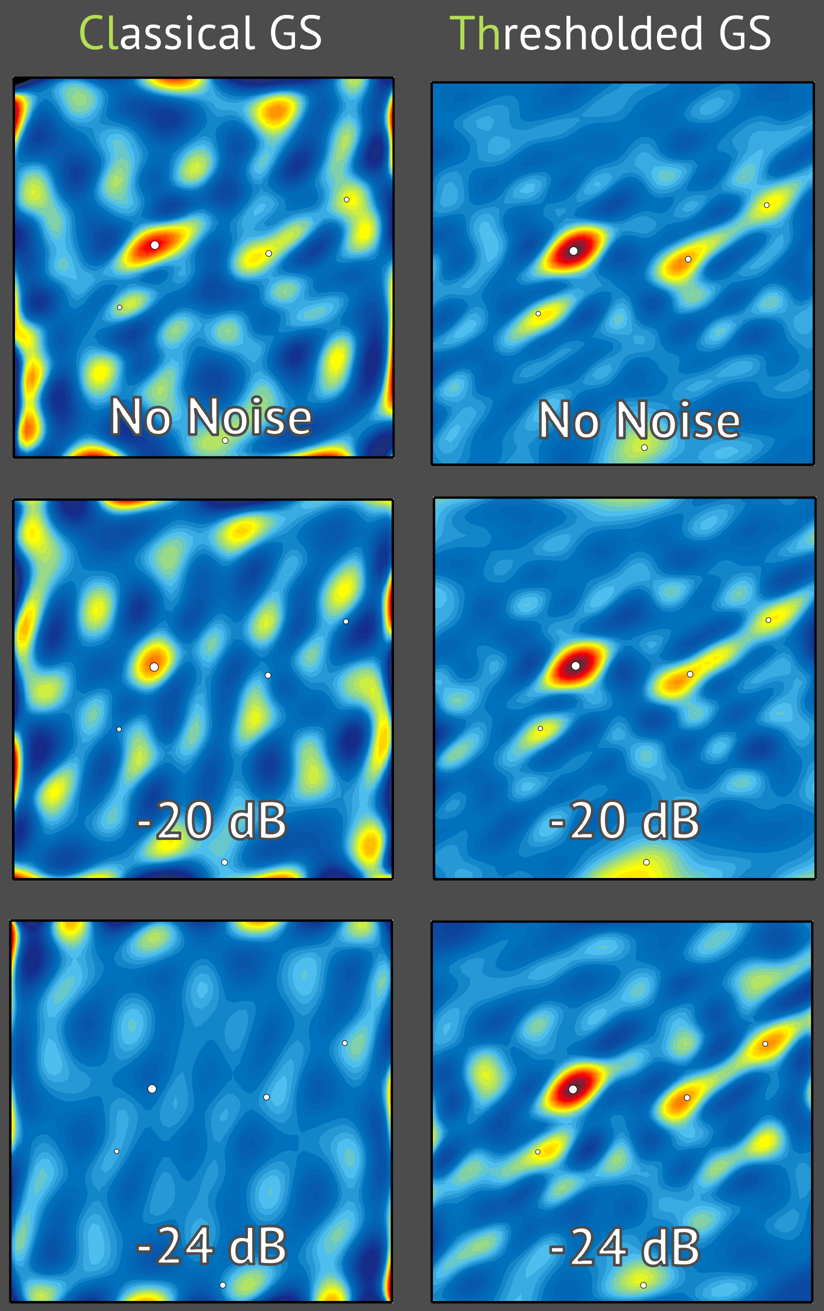
end

end

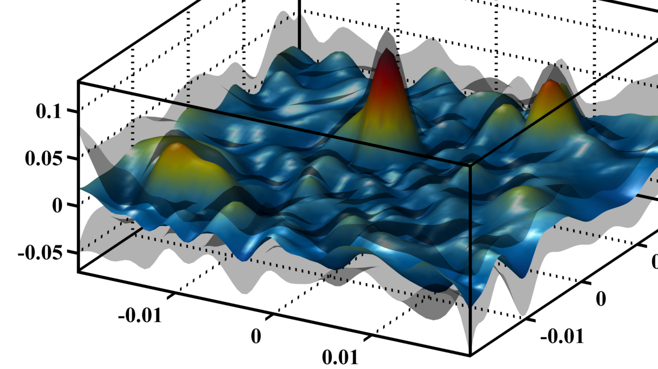
- For stability, apply thresholding



Merits of Thresholded Gram-Schmidt



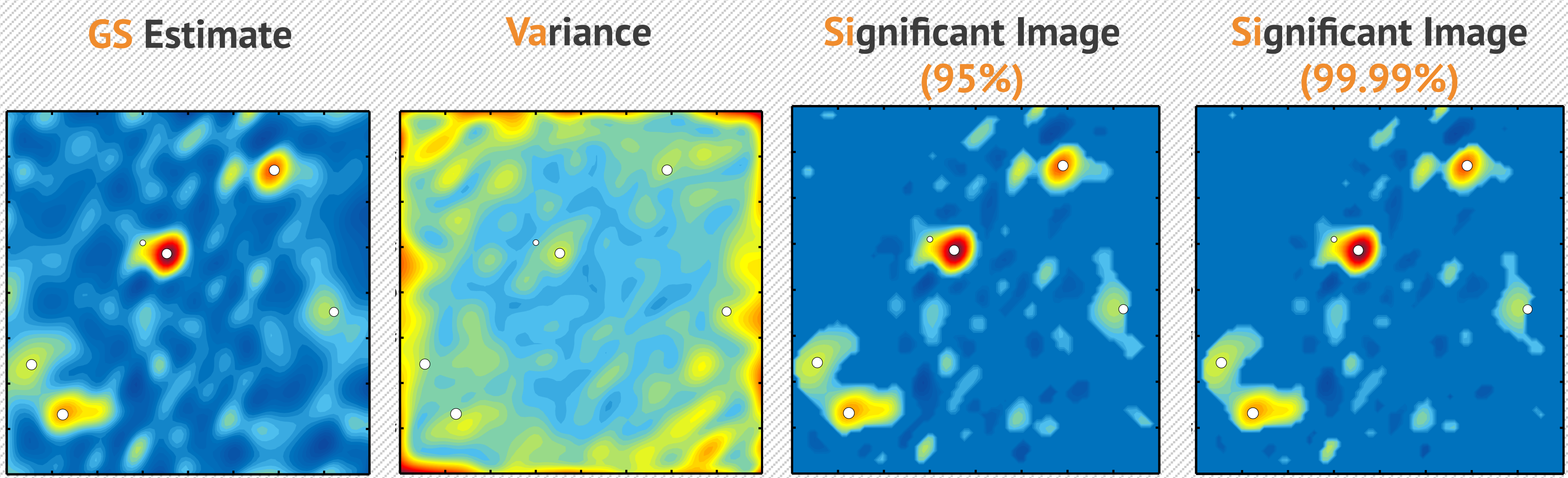
Statistical Testing of the GS Estimate



- In practice, visibilities are estimated. For Gaussian samples, we have a **Wishart distribution**.
- Gram-Schmidt estimate is linear on the data $\hat{\boldsymbol{I}}_{LS} = B_\perp^H \mathcal{G} \hat{\boldsymbol{V}}$. Hence,

$$\mathrm{Var}\left(\hat{\boldsymbol{I}}_{LS}\right) = \frac{1}{N_s} B_\perp^H \mathcal{G}(\Sigma \otimes \Sigma^*) \mathcal{G}^H B_\perp .$$

- Use asymptotic arguments to build global confidence intervals with the **Bonferroni method**.



Sparse Recovery

- Most of the sky is empty, we would like sparse estimates.
- Systematic approach: penalize the least squares problem. This yields the **LASSO** estimate, given by

$$\hat{\boldsymbol{I}}_{LASSO} = \mathrm{argmin}_{\boldsymbol{i} \in \mathbb{R}^{N^2}} \underbrace{\boxed{\|\boldsymbol{V} - B\boldsymbol{I}\|_2^2}}_{\text{Controls adequacy with the data}} + \underbrace{\boxed{\lambda \|\boldsymbol{I}\|_1 .}}_{\text{Controls sparsity of the estimate}}$$

- Very commonly used in **compressed sensing.**
- Less in radio interferometry: not the usual setup

Compressed Sensing: Nb of pixels << Nb of measurements
Radio Interferometry: Nb of pixels ≈ Nb of measurements

High noise

LASSO by Thresholding

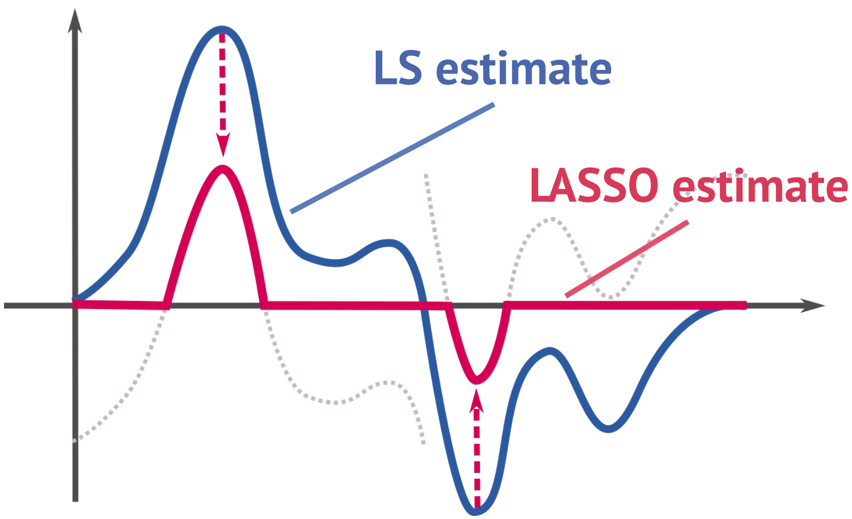
- Assume that the system has been orthogonalized.

$$\hat{\boldsymbol{I}}_{LASSO} = \operatorname{argmin}_{\boldsymbol{I} \in \mathbb{R}^{N^2}} \|\boldsymbol{V}_\perp - B_\perp \boldsymbol{I}\|_2^2 \; + \; \lambda \|\boldsymbol{I}\|_1.$$

- When $N^2 \leq J$, $B_\perp$ has orthogonal columns and hence

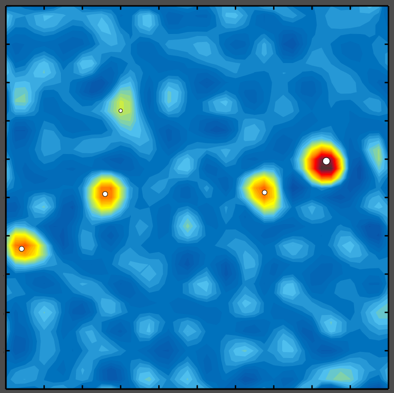
$$\hat{I}^i_{LASSO} = \operatorname{sgn}\left(\hat{I}^i_{LS}\right)\left(|\hat{I}^i_{LS}| - \frac{\lambda}{2}\right)^+.$$

Very Cheap!

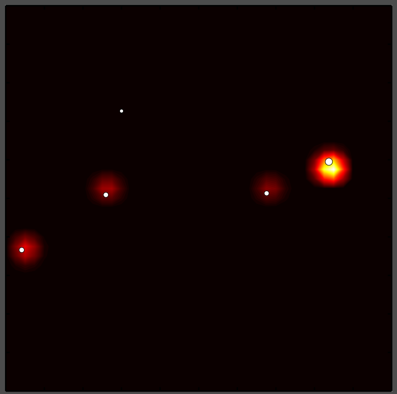


- Constraint on the resolution…

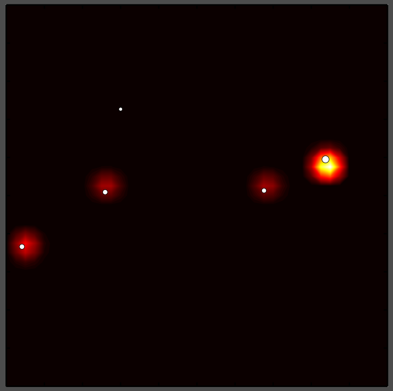
GS Estimate



LASSO from GS



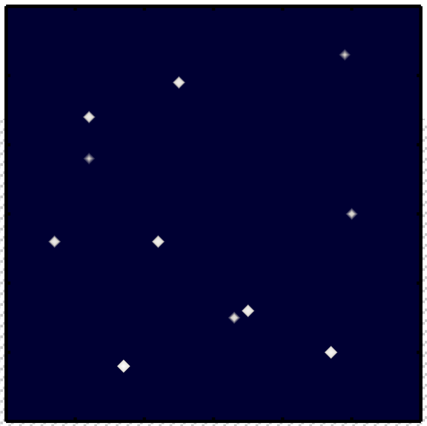
LASSO (FISTA)



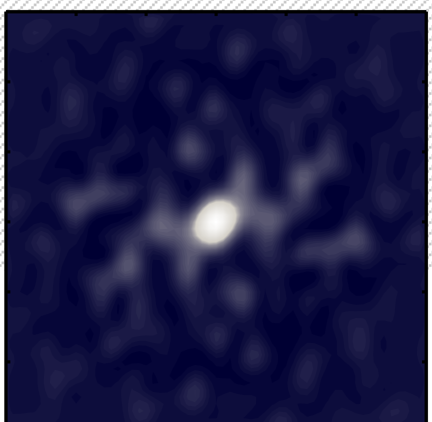
The Point Spread Function

- Response of the instrument to an **impulse signal** (single source in the sky). The point spread function is determined by the layout of the telescope.

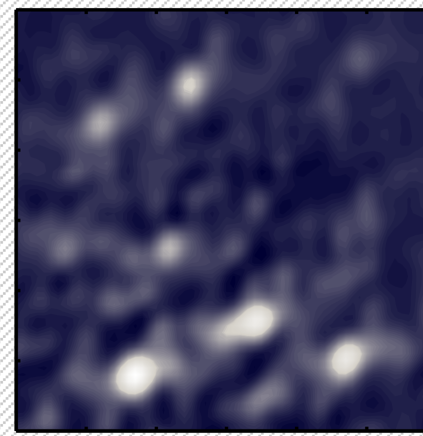
True Sky



PSF of the telescope



Sky as seen by the telescope

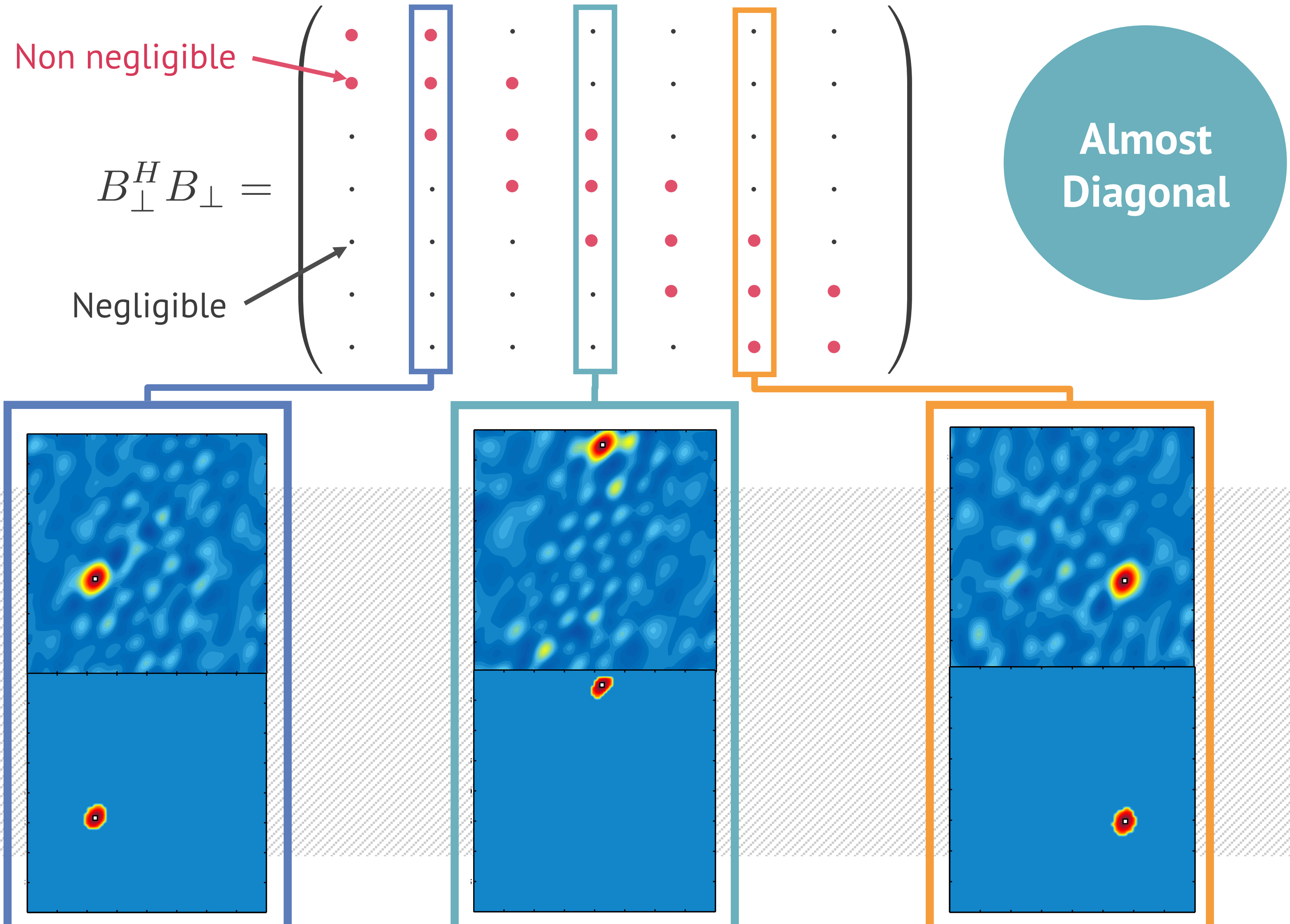


Interferometers layouts are chosen to optimize the Point Spread Function

Gram-Schmidt and the PSF

- Assume that $B_\perp$ have been obtained with the Gram-Schmidt procedure. Then, we can show that

$$B_\perp^H B_\perp =$$



PSF for pixel *i*

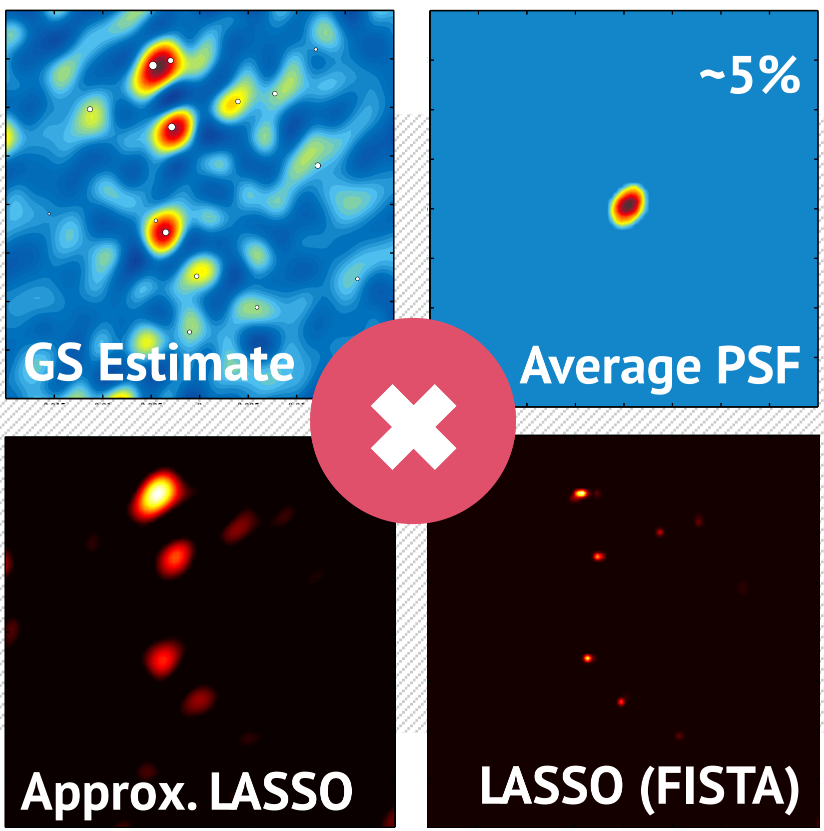
Effective Support

Approximate LASSO from Gram-Schmidt

- If $B_\perp^H B_\perp \simeq \mathrm{diag}(B_\perp^H B_\perp)$, then we can approximate the LASSO

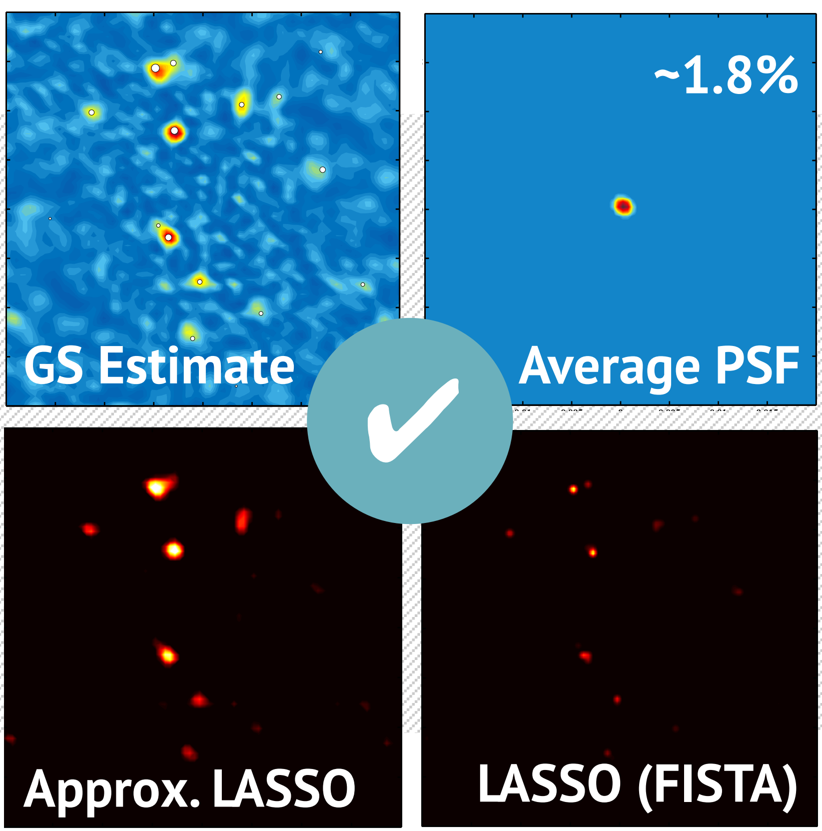
$$\hat{I}^i_{LASSO} \simeq \frac{\mathrm{sgn}\left(\hat{I}^i_{LS}\right)}{\mu_i}\left(\left|\hat{I}^i_{LS}\right| - \frac{\lambda}{2}\right)^+ .$$

12 LOFAR stations

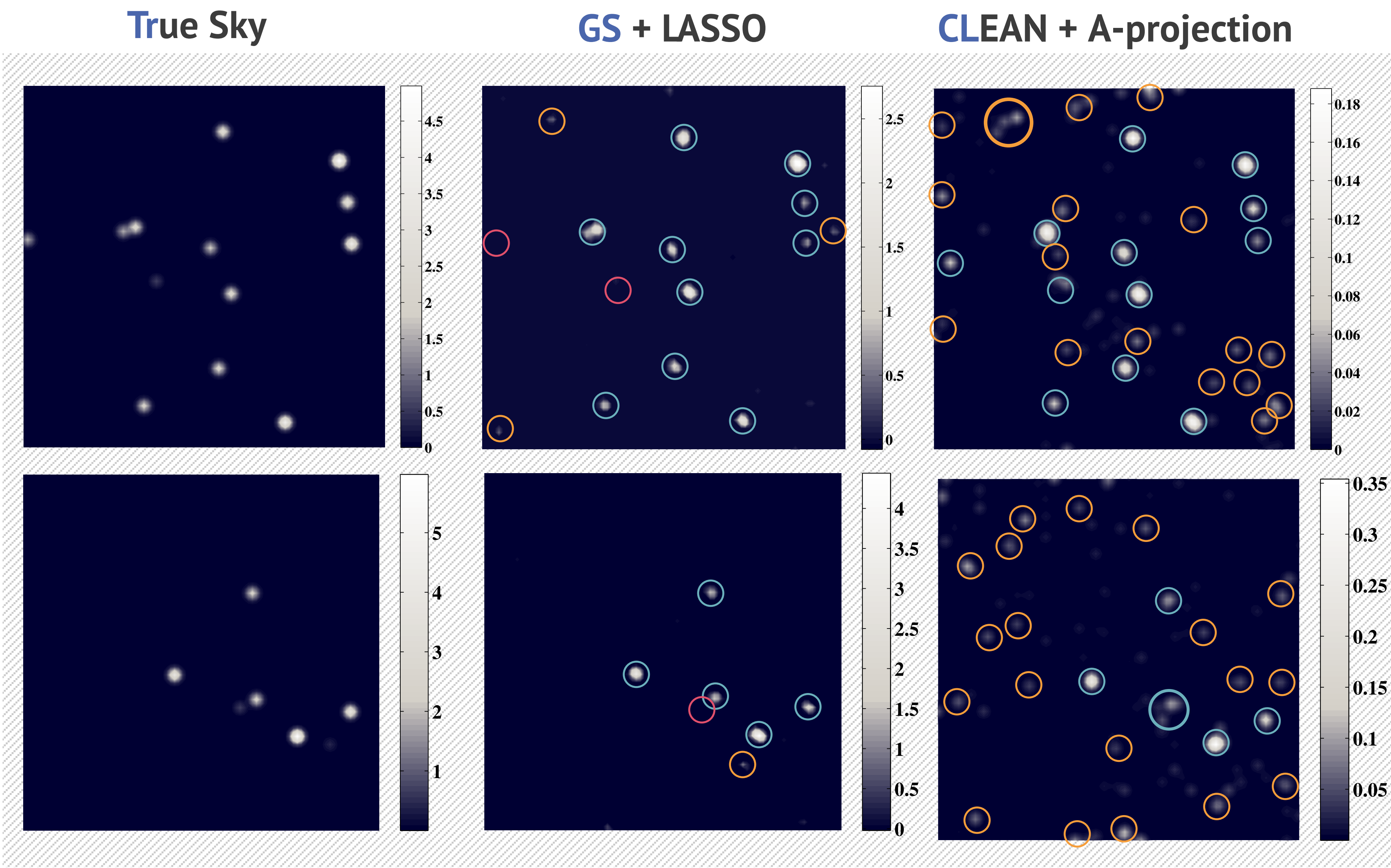


VS.

24 LOFAR stations

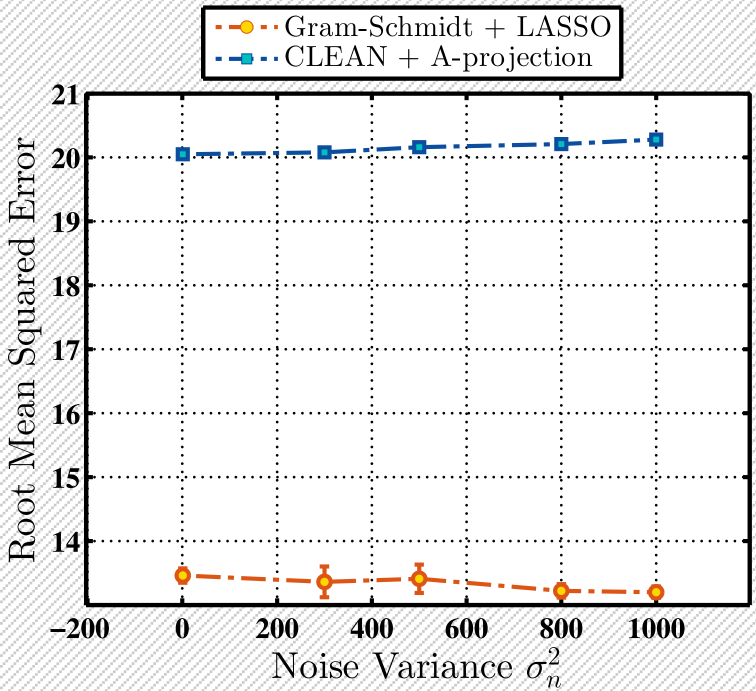


Comparison with CLEAN

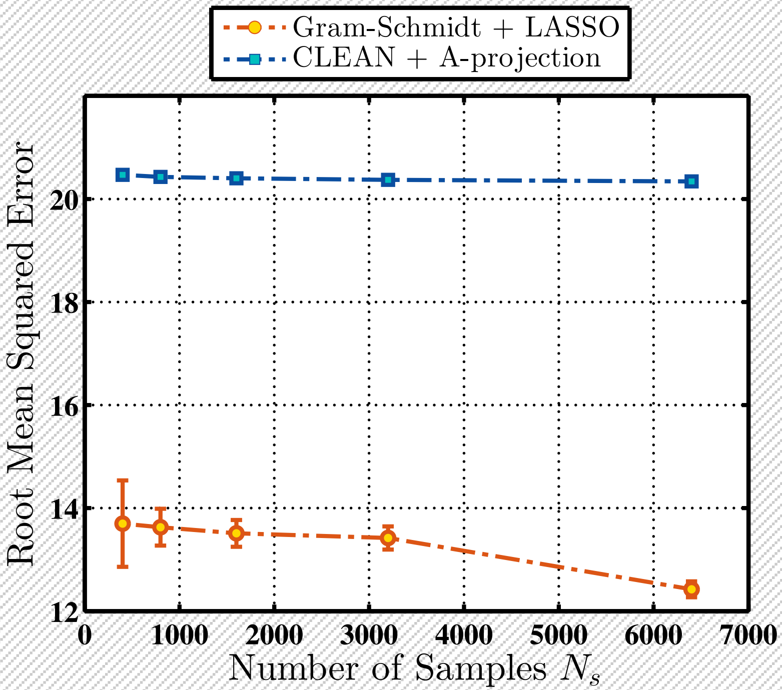


Comparative Sensitivity Analysis

- For a given sky, we compared the sensitivity of GS+LASSO and CLEAN+A-projection

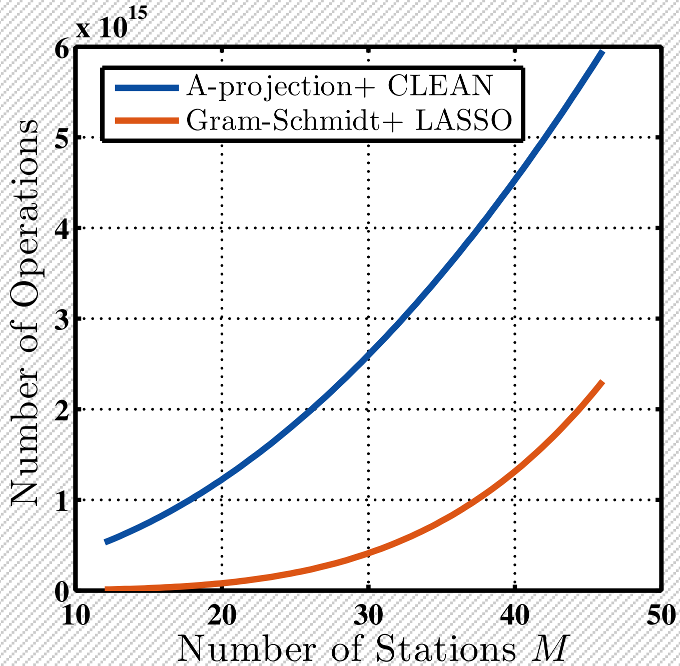


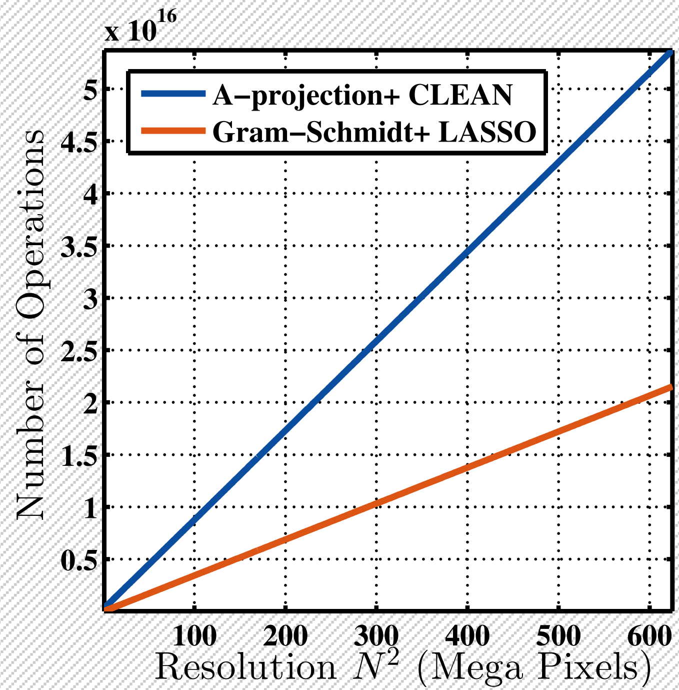
GS+LASSO IMAGER MORE ROBUST TO THE NOISE



GS+LASSO IMAGER ACCURACY INCREASES WITH NB OF SAMPLES

Complexity Analysis (LOFAR)

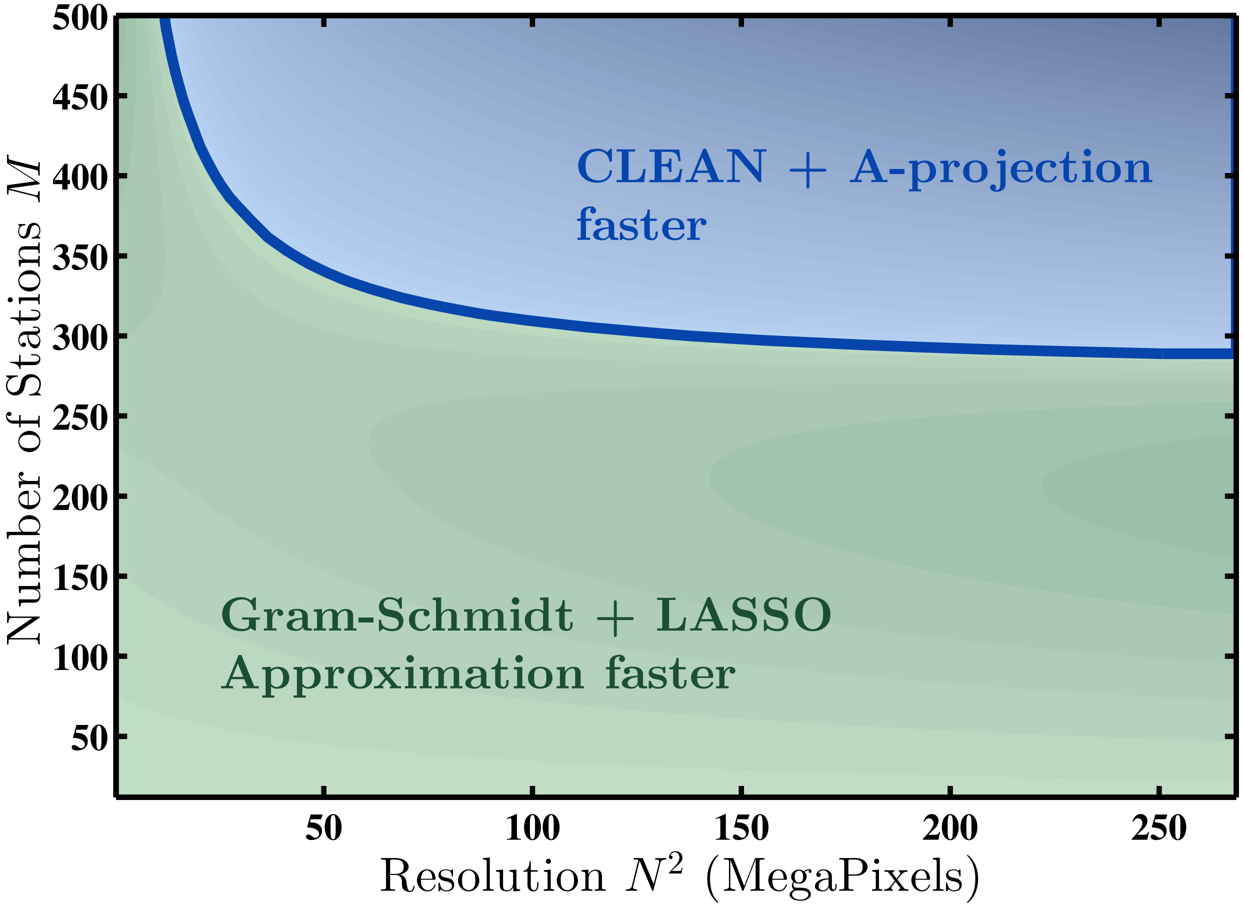




c_{Aproj}/c_{GS}		Number of Stations		
		$M = 24$	$M = 38$	$M = 46$
Resolution	$N^2 = 1024 \times 1024$	34.051	13.364	9.0748
	$N^2 = 2048 \times 2048$	15.804	8.6764	4.1174
	$N^2 = 4096 \times 4096$	11.304	4.2796	2.8832
	$N^2 = 8192 \times 8192$	10.239	3.8369	2.579

2 to 34 times faster

Complexity Analysis (SKA)



GS+LASSO IMAGER IS FASTER FOR MANY SKA PRACTICAL CASES

Conclusions

- Beamforming breaks the intimate relationship with Fourier domain: cannot interpret visibilities as *uv*-samples.
- Performing a **QR-decomposition of the system** results in a more *intuitive*, *natural* and *flexible* imaging pipeline.
- Even though at a very early stage, the Gram-Schmidt imaging pipeline is **more accurate** and **faster** than state-of-the-art for LOFAR and many SKA scenarios.
- QR-decomposition is currently oversampled. There is room for improvement.
- Redundancies exist between different time intervals and frequency channels. We believe our framework is capable of exploiting such redundancies.

Contributions